1st Year exam - Andres Rivero Gamez

Andres Rivero Gamez

2022-06-13

Name: Andres Rivero Gamez

PID: A59010594

First, uploading the packages that we will need for this.

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Then, reading the information I downloaded from the website onto R.

```
CovidData<-read.csv("covid19.csv")
```

Using the lubridate package to transform the date column into the right format.

```
CovidData$date<-ymd(CovidData$date)
```

Deleting the "Total" values from the data frame.

```
CovidData1<- CovidData[CovidData$variant_name !="Total", ]
```

Plotting the data.

(And spending more time than I want to admit making it look exactly like the exam page one)

```
ggplot(data = CovidData1) + aes(x = date, y= percentage, col=variant_name) +
  ggtitle("Covid-19 Variants in California") +
  labs(x="", y="Percentage of sequenced specimenes",
       caption = "Data source: <https://www.cdph.ca.gov>")+
  scale_x_date(breaks= "months", date_labels = "%b %Y") + theme_bw() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1),legend.title=element_blank()) +
  geom_line() + geom_point(size=.1)
```